

Keyboards are an old-fashioned way of entering data. **Simon Handby** explains how modern handwriting-recognition software makes computers easier to use

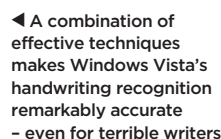
Handwriting recognition has found its way on to full-sized computers, too. The tablet PC version of Windows XP and all versions of Windows Vista are able to accept handwritten input. Here's we'll explain how it works.

Before a computer can recognise written words, it needs to be able to see them. This may be achieved by scanning in a document, which the computer then analyses in its entirety. The optical character recognition (OCR) performed

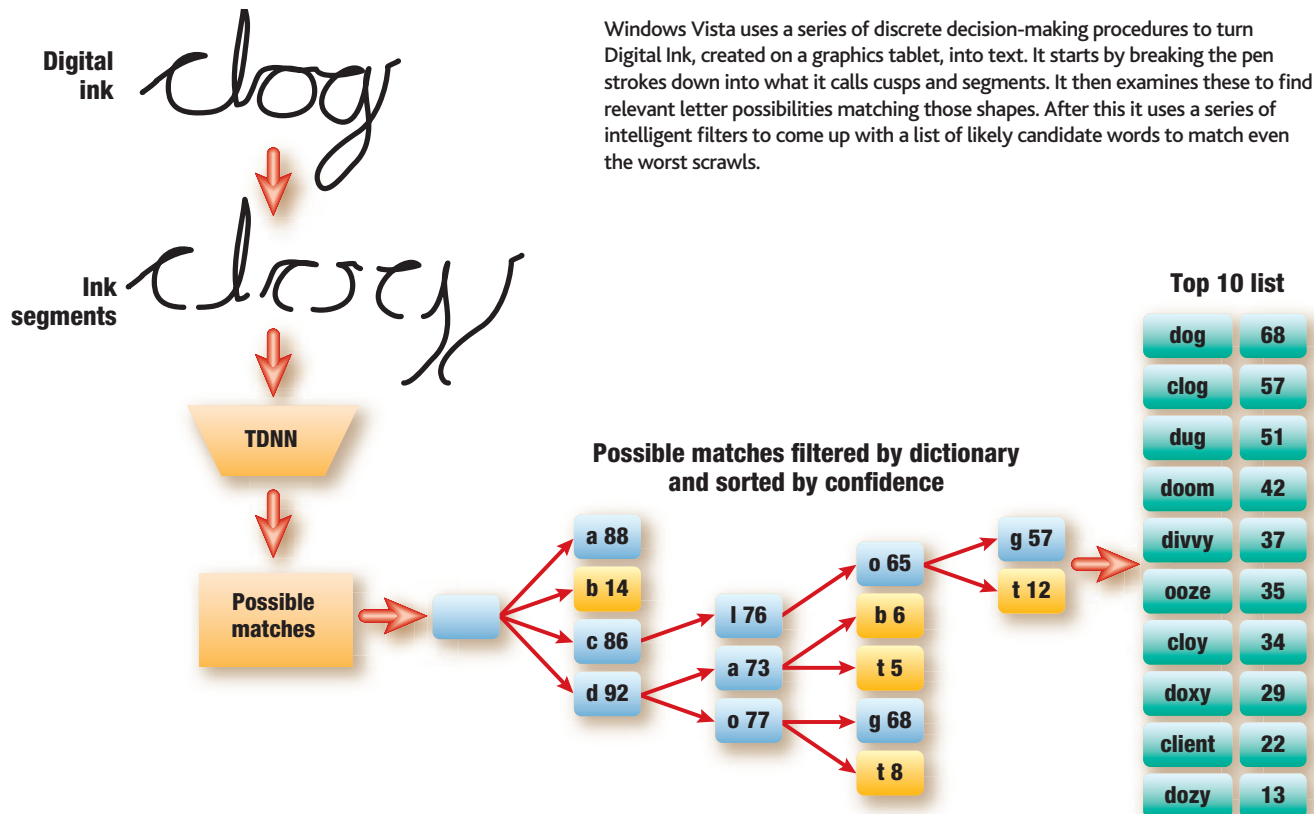
The analysis of words contained in digital ink varies depending on the platform, but it generally involves a preparation stage, then a recognition stage, followed by a series of refinements and filters that improve the accuracy of the final result. Windows Vista has an up-to-date, effective and widely distributed recognition engine, so we'll examine it in detail.

Once the strokes have been reordered in this way, they are passed into the text recogniser, which begins to analyse their content. The first stage involves breaking the ink into segments using a simple algorithm, which slices ink apart vertically in places that might form the bottom of a letter.

Recognisers are specific to a language, with regards to both characters and words. For example, a French recogniser uses a French



Character analysis Understanding Vista's handwriting recognition



dictionary, and will consider a grave accent above an A, E or U among its possible outcomes. An English recogniser won't include such accents when considering possible segment matches.

UNDERSTANDING FACTOIDS

Using the confidence scores that the TDNN assigns to each of the possibilities it identifies, the recogniser can rank all of the remaining valid outcomes based on the combined confidence score of their parts – in effect finding a path of least resistance from start to end of the digital ink. However, given more information it may be possible to filter out further invalid results, or to place a better match at the top of the list.

Windows Vista's handwriting recognition refines its results further through the use of 'factoids'. These are filters that can be applied by software developers to tell the recogniser what type of input to expect.

For example, a UK postcode factoid instructs the recogniser to look for input in the pattern of a postcode: one or two capital letters, followed by one or two numbers, a space, one number and two more capital letters. Without the factoid, the recogniser would be extremely unlikely to identify such a specific format, but factoids need to be used with caution: applying a postcode factoid to a standard text field would produce disastrously inaccurate results.

A final way that Vista improves the accuracy of its handwriting recognition is by allowing customisation. On a simple level, this means that the system includes words from the custom dictionary among the possibilities during recognition, allowing it to learn and recognise the names or technical terms you use that it may not know by default.

There are also forms of personalisation available to help the system learn more about your writing style and thus improve its accuracy. Vista's opt-in Automatic Learning feature is an ongoing process that allows the recogniser to store information about the words used and how they are written, while the Handwriting Recognition Personalization tool prompts you for a set of handwriting samples from which it can improve its understanding of your writing. Alternatively, if a specific word is causing repeated problems, you can choose to train the system to recognise it correctly. If you buy a new PC or need to re-install an existing one, you can use Windows Easy Transfer to copy across everything Vista has learned about your writing.

SPOT THE MISTAKE

After applying several layers of analysis and filtering, modern handwriting-recognition software is remarkably accurate, but it isn't foolproof. Once the software has identified the most probable input, the interface normally makes it easy for the user to replace the system's best match with one of the next most likely alternatives. This ability to correct mistakes is one great advantage that online recognition has over offline processing.

Some security systems actually rely on the difficulty experienced by machines in recognising text, and the comparative ease with which humans can do it. Anyone who often comments on blog posts or who has signed up to a new online email account is likely to have experienced a Human Interactive Proof (HIP). This is a test to establish that the user is a real human and not a machine employed by a spammer.

Commonly used tests known as CAPTCHAs (Completely Automated Public Turing tests to

tell Computers and Humans Apart) pose a question that can be quickly and accurately answered by a human but flummox a computer – the most common form being where the user must recognise some obscured text. The reCAPTCHA project uses words that have defeated OCR software during the digitisation of older books within its visitor verification procedures.

UNDERSTANDING FACTOIDS

A very specialised type of handwriting recognition is used in biometrics. Here, the objective is to recognise a specific person's handwriting – usually a signature – rather than the content within it. The procedure for this differs from regular handwriting recognition; it doesn't just examine the shape of a signature, but also considers other data such as the speed of the handwritten strokes, the pressure applied to the stylus and the angle at which it is held. By recording and testing for this additional data, the system can resist simple fraudulent attempts to 'trace' a signature's outline on to a tablet.

FURTHER INFORMATION

Wikipedia entry on handwriting recognition
http://en.wikipedia.org/wiki/Handwriting_recognition

Wikipedia entry on neural networks http://en.wikipedia.org/wiki/Artificial_neural_network

Developer guide to handwriting recognition on Tablet PCs www.tinyurl.com/36rwp

ReCAPTCHA book digitisation project
<http://recaptcha.net>

Recognition expert Dr Venu Govindaraju
<http://www.cubs.buffalo.edu/govind>